Mining associations and roles: role of feature extraction

Goran Nenadic
School of Computer Science
University of Manchester
G.Nenadic@manchester.ac.uk

One of the ultimate aims of biomedical text mining would be to extract both explicit and implicit associations between different types of entities. In addition, assigning roles that entities have or may have in biological processes is also of interest. In this talk I will be discussing our experience in selecting and engineering textual features that can help in mining associations and roles from literature. Depending on tasks and entities involved, we have used four types of features: from simple words and terms, to words and semantic classes, to textual contexts, to contexts augmented with additional background attributes. The main epilogue is that both NLP- and domain-knowledge driven feature engineering are needed for successful mining of associations and roles.